ON THE MARKOV EQUIVALENCE OF CHAIN GRAPHS, UNDIRECTED GRAPHS, AND ACYCLIC DIGRAPHS

by

Steen A. Andersson
David Madigan
Michael D. Perlman

TECHNICAL REPORT No. 281
December 1994

Department of Statistics, GN-22
University of Washington
Seattle, Washington 98195 USA
On the Markov Equivalence of Chain Graphs, Undirected Graphs, and Acyclic Digraphs*

Steen A. Andersson, David Madigan, and Michael D. Perlman†
Indiana University and University of Washington
December 28 1994

Abstract

Graphical Markov models use undirected graphs (UDGs), acyclic directed graphs (ADGs), or (mixed) chain graphs to represent possible dependencies among random variables in a multivariate distribution. Whereas a UDG is uniquely determined by its associated Markov model, this is not true for ADGs or for general chain graphs (which include both UDGs and ADGs as special cases). This paper addresses three questions regarding the equivalence of graphical Markov models: when is a given chain graph Markov equivalent (1) to some UDG? (2) to some (at least one) ADG? (3) to some decomposable UDG? The answers are obtained by means of an extension of Frydenberg's (1990) elegant graph-theoretic characterization of the Markov equivalence of chain graphs.

1 Introduction

The use of graphs to represent dependence relations among random variables, first introduced by Wright (1921), has generated considerable research activity, especially since the early 1980s. Particular attention has been devoted to graphical Markov models, where a graph is used to specify certain conditional independence relations among the variables. These models have found diverse applications, including image analysis, spatial statistics, categorical data analysis, pedigree analysis, and expert systems. The recent books by Pearl (1988), Whittaker

---

*Research supported in part by the U.S. National Science Foundation and the U.S. National Security Agency.
†Address for correspondence: Department of Statistics, GN-22, University of Washington, Seattle WA 98195, USA. (michael@stat.washington.edu)
In these graphical models, the vertices of the graph represent random variables and the links between the vertices (which can be either directed or undirected) represent the absence of conditional independence. Much of the research on graphical models in the 1980s focused on “pure” graphs, i.e., either undirected graphs, where all the links are undirected, or acyclic digraphs where all the links are directed and no directed cycles occur. For example, the acyclic digraph (ADG) of Figure 1(a) embodies the assumption that \( \beta \) and \( \gamma \) are conditionally independent given \( \alpha \). The undirected graph (UDG) of Figure 1(b) represents the same assumption, although in general, directed and undirected graphs with the same vertices and links will represent different sets of conditional independencies.

![Diagram](image)

*(a) A simple acyclic digraph (ADG) and (b) a simple undirected graph (UDG).*

In the mid-1980s, Lauritzen, Wermuth, and Frydenberg introduced graphical models defined by “chain graphs” (Lauritzen and Wermuth, 1989, Frydenberg, 1990). This followed earlier work in this direction by Goodman (1973), Asmussen and Edwards (1983), and Kiiveri et al. (1984). Chain graphs may have both directed and undirected edges, and include both the ADGs and UDGs as special cases. The defining property of a chain graph is that it contains no cycles involving one or more directed edges. Chain graphs provide much of the focus for current research on modelling statistical dependence, see for example, Lauritzen.

---

1We now use the phrase “acyclic digraph” rather than the more common (but inaccurate) “directed acyclic graph” after Brian Alspach kindly pointed out to us that these adjectives do not commute.
(1989), Wermuth and Lauritzen (1990), and Cox and Wermuth (1993).

While chain graphs do provide a more general modeling class, ADG models admit especially simple statistical analysis, and have become popular across a diverse range of applications; see, for example, Spiegelhalter et. al. (1993), Neapolitan (1990), Pearl (1988), Spiegelhalter and Lauritzen (1990), Lauritzen and Spiegelhalter (1988), York et al. (1995), and Madigan and Raftery (1994). In particular, ADG models provide a convenient recursive factorization of the joint probability density function (see (3.1)), an elegant framework for Bayesian analysis, and, in expert system applications, a simple causal interpretation. In the multinomial and multivariate normal cases, the likelihood function (i.e., both the joint probability density function and the parameter space) factorizes and admits explicit maximum likelihood estimates. Furthermore, the only UDGs that provide these conveniences are the decomposable UDG models; these decomposable UDGs are exactly those that are Markov equivalent to (i.e., have the same Markov properties as) some ADG (see Figure 2 and Corollary 4.5 below).

Whereas a UDG is uniquely determined by its associated Markov model, this need not be true for ADGs or for chain graphs. In particular, a given chain graph $G$ may be Markov equivalent to some UDG, to some ADG(s), or to both. In particular, if $G$ is determined to be equivalent to an ADG, then its statistical analysis can be simplified substantially.

These considerations suggest the following three questions:

1. When is a chain graph, $G$, Markov equivalent to some (necessarily unique) UDG?

2. When is a chain graph, $G$, Markov equivalent to some (not necessarily unique) ADG?
   In this case, is there an efficient algorithm to generate at least one (or all) Markov equivalent ADGs from $G$?

3. When is a chain graph, $G$, Markov equivalent to some (necessarily unique) decomposable UDG?

Question 1 was essentially answered by Frydenberg (1990, p.348). Proposition 8.2 of Lauritzen and Wermuth (1989) presents a sufficient condition for a given chain graph, $G$, to be Markov equivalent to a given ADG. Further, they conjecture that their condition is also necessary. Their question is related to, but not identical to, our Question 2. Similarly, their Proposition 8.3 addresses a question related to our Question 3.

In this paper we resolve Questions 1, 2, and 3 in a unified fashion, in Propositions 4.1, 4.2, and 4.3, respectively. Our main result is Proposition 4.2, where we provide necessary
and sufficient conditions for a given chain graph, $G$, to be Markov equivalent to some ADG. Also, we present the required algorithm for generating at least one Markov equivalent ADG from $G$. (See Remark 4.1 regarding how to generate all Markov equivalent ADGs from $G$.) This work is based on our extension to general probability measures (see Theorem 3.1) of Frydenberg's (1990) elegant graph-theoretic characterization of the Markov equivalence of two chain graphs. Thus our arguments are graph-theoretic rather than probabilistic.

Various authors have shown (with varying degrees of generality) that the intersection of the classes of UDG models and ADG models is the class of decomposable UDG models (Wermuth (1980, Proposition 5), Wermuth and Lauritzen (1983, Proposition 5), Asmussen and Edwards (1983, Corollary 3.4), and Kiiveri, Speed, and Carlin (1984, Corollary, p.39). By combining our answers to Questions 1, 2, and 3, we obtain a purely graph-theoretic demonstration of this result in complete generality (Corollary 4.5).

![Figure 2: Four classes of graphical Markov models. Chain graphs include UDGs, ADGs, and decomposable UDGs as special cases.](image)

### 2 Definitions and Notation

Our development closely follows that of Frydenberg (1990), with one exception noted below. We consider multivariate probability distributions on a product probability space...
\[ X \equiv \times_{\alpha \in V} X_{\alpha} \], where \( V \) is a finite index set and each \( X_{\alpha} \) is sufficiently regular to ensure the existence of regular conditional probabilities. Such distributions are conveniently represented by a random variate \( X := (X_{\alpha} : \alpha \in V) \in X \). For any subset \( A \subseteq V \), we define \( X_A := (X_{\alpha} : \alpha \in A) \). We often abbreviate \( X_{\alpha} \) and \( X_A \) by \( \alpha \) and \( A \), respectively, and define \( X_A \equiv \text{constant} \).

A graphical model is defined by a collection of conditional independencies among the component random variates \((X_{\alpha} : \alpha \in V)\), which collection is represented by a graph \( G \equiv (V, E) \) with vertex set \( V \). The set of edges, \( E \), is a subset of \( E^*(V) \equiv (V \times V) \setminus \{(\alpha, \alpha) : \alpha \in V\} \), i.e., a set of ordered pairs of distinct vertices. An edge \((\alpha, \beta) \in E \) whose opposite \((\beta, \alpha) \in E \) also, is called an undirected edge and appears as a line \( \alpha \rightarrow \beta \) in our figures, whereas an edge \((\alpha, \beta) \in E \) whose opposite \((\beta, \alpha) \notin E \) is called a directed edge and appears as an arrow: \( \alpha \rightarrow \beta \). (Our notation differs from Frydenberg’s in this regard: he uses the notation \( C \rightarrow \) instead of \( \alpha \rightarrow \beta \) in his text, although not in his figures.)

If \( A \subseteq V \) is a subset of the vertex set, it induces a subgraph \( G_A = (A, E_A) \), where the edge set \( E_A \equiv E \cap (A \times A) \) is obtained from \( G \) by retaining all edges with both endpoints in \( A \). If \( A \subseteq B \subseteq V \), clearly \( G_A = (G_B)_A \). A graph is complete if all pairs of vertices are joined by an edge. Trivially, the empty graph is complete. A subset \( A \subseteq V \) is complete in \( G \) if it induces a complete subgraph. A complete subset that is maximal with respect to inclusion is called a clique.

For a graph \( G = (V, E) \), we will denote the skeleton of \( G \) by \( G^u = (V, E^u) \), where \( E^u = \{(\alpha, \beta) : (\alpha, \beta) \in E \text{ or } (\beta, \alpha) \in E \} \); \( G^u \) is just the underlying undirected graph associated with \( G \). For any subset \( A \subseteq V \), \((G_A)^u = (G^u)_A \). Two vertices \( \alpha \) and \( \beta \in V \) are called adjacent in \( G \) if \((\alpha, \beta) \in G^u \).

For a subset \( A \subseteq V \), \( \text{bd}(A) \equiv \{ \beta \in V \setminus A \mid (\beta, \alpha) \in E \text{ for some } \alpha \in A \} \) and \( \text{cl}(A) \equiv \text{bd}(A) \cup A \) denote the boundary and closure of \( A \) in \( G \), respectively. The neighbors of \( \alpha \) in \( G \), denoted by \( \text{nb}(\alpha) \equiv \{ \beta \in V \setminus \alpha : (\alpha, \beta) \in E \text{ and } (\beta, \alpha) \in E \} \), are those vertices linked to \( \alpha \) by undirected edges. The parents of \( \alpha \) in \( G \), denoted by \( \text{pa}(\alpha) \equiv \{ \beta \in V \setminus \alpha : (\beta, \alpha) \in E \text{ and } (\alpha, \beta) \notin E \} \), are those vertices linked to \( \alpha \) by directed edges. For subsets \( A \subseteq B \subseteq V \), we write \( \text{bd}_B(A), \text{cl}_B(A), \text{pa}_B(A) \), etc., to denote the boundary, closure, parents, etc., of \( A \) in the induced subgraph \( G_B \).

A path \( \pi \) of length \( n \geq 1 \) from \( \alpha \) to \( \beta \) in \( G \) is a sequence \( \pi \equiv \{\alpha_0, \alpha_1, \ldots, \alpha_n\} \subseteq V \) of distinct vertices such that \( \alpha_0 = \alpha, \alpha_n = \beta, \) and \((\alpha_{i-1}, \alpha_i) \in E \) for all \( i = 1, \ldots, n \). If \((\alpha_{i-1}, \alpha_i) \) is directed for at least one \( i \), we call the path directed, and if this is not the case, we call it undirected. A cycle is a path with the modification that \( \alpha_n = \alpha_0 \). A graph is called
a chain graph if it does not contain any directed cycles. If the graph has only undirected edges it is an undirected graph (UDG). If all the edges are directed, and the graph contains no directed cycles, the graph is said to be an acyclic digraph (ADG).

Let \( G \equiv (V, E) \) be a UDG. A set of vertices \( A \subseteq V \) is connected in \( G \) if, for every distinct \( \alpha, \beta \in A \), there is a path from \( \alpha \) to \( \beta \). For pairwise disjoint subsets \( A(\neq \emptyset), B(\neq \emptyset) \), and \( S \) of \( V \), \( A \) and \( B \) are separated by \( S \) in \( G \) if all paths within \( G \) from vertices in \( A \) to vertices in \( B \) include vertices in \( S \). Note that if \( S = \emptyset \), then \( A \) and \( B \) and separated by \( S \) in \( G \) if and only \( A \) and \( B \) are not connected in \( G \). Further, if \( A \) and \( B \) are not connected in \( G \), then they are separated by any subset \( S \) disjoint from \( A \) and \( B \).

A subset \( A \subseteq V \) is simplicial in the UDG, \( G \), if its boundary is complete. A pair \((A, B)\) of nonempty subsets of \( V \) is said to form a decomposition of \( G \) if \( V = A \cup B \), \( A \cap B \) is complete, and \( A \cap B \) separates \( A \setminus B \) from \( B \setminus A \) in \( G \). When this is the case we say that \((A, B)\) decomposes \( G \) into the components \( G_A \) and \( G_B \). If the sets \( A \) and \( B \) in \((A, B)\) are both proper subsets of \( V \), the decomposition is proper. An undirected graph is said to be decomposable if it is complete, or if there exists a proper decomposition \((A, B)\) into decomposable subgraphs \( G_A \) and \( G_B \).

A UDG is chordal if every cycle of length \( n \geq 4 \) possesses a chord, that is, two nonconsecutive vertices that are neighbors. A well known result states that a UDG is decomposable if and only if it is chordal, cf., Lauritzen, Speed, and Vijayan (1984, Theorem 2) or Whittaker (1990, Proposition 12.4.2).

For the remainder of this section, let \( G \equiv (V, E) \) be a chain graph. If \( A \subseteq V \) and \( \alpha, \beta \in A \), we write \( \alpha \leq_A \beta \) if \( \alpha = \beta \) or there is a path from \( \alpha \) to \( \beta \) in \( G_A \). If both \( \alpha \leq_A \beta \) and \( \beta \leq_A \alpha \), then we write \( \alpha \approx_A \beta \). When \( A = V \), we write \( \leq (\equiv \leq_G) \) for \( \leq_V \) and \( \approx (\equiv \approx_G) \) for \( \approx_V \). Frydenberg (1990) notes that \( \approx \) is an equivalence relation on \( V \) and denotes the induced set of equivalence classes by \( \mathcal{T}(G) \), the set of chain components of \( G \). Equivalently, \( \mathcal{T}(G) \) is the set of connected components of the undirected graph obtained from \( G \) by removing all directed edges. Note that for each \( \tau \in \mathcal{T}(G) \), \( \text{bd}(\tau) = \text{pa}(\tau) \equiv \{ \text{pa}(\alpha) \mid \alpha \in \tau \} \). Both UDGs and ADGs are special cases of chain graphs: a UDG has only a single chain component, while for an ADG, every chain component consists of a single vertex.

For any subset \( A \subseteq V \), the induced subgraph \( G_A \) is also a chain graph, and \( \mathcal{T}(G_A) = \{ \tau \cap A \mid \tau \in \mathcal{T}(G) \} \). Thus, if \( \tau \in \mathcal{T}(G) \) and \( \tau \subseteq A \), then \( \tau \in \mathcal{T}(G_A) \).

Let \( \alpha < \beta \) stand for the statement "there exists a directed path from \( \alpha \) to \( \beta \)." We define the future of \( \alpha \) in \( G \) by \( \phi(\alpha) = \{ \beta \mid \alpha < \beta \} \), and the past of \( \alpha \) in \( G \) by \( \pi(\alpha) = \{ \beta \mid \beta < \alpha \} \). If \( \tau \in \mathcal{T}(G) \), the future or the past is the same for all vertices in \( \tau \) and thus we may use
\( \phi(\tau) \) and \( \pi(\tau) \) to denote their common future and past in \( G \), respectively. Furthermore, we call \( \tau \) terminal in \( G \) if \( \phi(\tau) \) is empty and initial in \( G \) if \( \pi(\tau) \) is empty.

A subset \( A \subseteq V \) is called a \( G \)-anterior set (or simply an anterior set if \( G \) is understood) if it can be generated by stepwise removal of terminal chain components. (Note that the removal of a terminal chain component might render other chain components terminal in the remaining graph.) Equivalently, it can be shown that \( A \) is anterior if and only if \( \beta \in A \) whenever \( \alpha \in A, \beta \in V, \) and \( \beta \leq \alpha \). Also equivalently, \( A \) is anterior if and only if \( \text{bd}(A) = \emptyset \).

If \( A \) is anterior and \( \tau \) is a terminal chain component of \( G \) such that \( \tau \subseteq A \), then \( \tau \) is a terminal chain component of \( G_A \).

If \( A \) and \( B \) are anterior sets, then \( A \cap B \) and \( A \cup B \) are anterior. If \( B \subseteq A \subseteq V \) and \( B \) is \( G \)-anterior, then \( B \) is \( G_A \)-anterior. If \( B \subseteq A \subseteq V \) and \( A \) is \( G \)-anterior, then \( B \) is \( G \)-anterior if and only if \( B \) is \( G_A \)-anterior.

For any subsets \( B \subseteq A \subseteq V \), let \( \text{an}_A(B) \) denote the smallest \( G_A \)-anterior subset of \( A \) that contains \( B \). Clearly \( \text{an}_A(B) = \{ \alpha \in A \mid \alpha \leq_A \beta \text{ for some } \beta \in B \} \), so if \( A \) is \( G \)-anterior and \( B \subseteq A \), then \( \text{an}_A(B) = \text{an}(B) \). Note that for any subsets \( B \subseteq A \) and \( C \subseteq A \), \( \text{an}_A(B \cup C) = \text{an}_A(B) \cup \text{an}_A(C) \) (but \( \cup \) cannot be replaced by \( \cap \)).

We call a triple \( (\alpha, B, \beta) \) a complex in \( G \) if \( B \) is a connected subset of a chain component \( \tau \in T(G) \) and \( \alpha \) and \( \beta \) are two non-adjacent vertices in \( \text{bd}(\tau) \cap \text{bd}(B) \). Further, we call \( (\alpha, B, \beta) \) a minimal complex in \( G \) if \( B = B' \) whenever \( B' \) is a subset of \( B \) and \( (\alpha, B', \beta) \) is a complex. Every complex \( (\alpha, B, \beta) \) contains at least one minimal complex, \( (\alpha, B', \beta), B' \subseteq B \). A minimal complex \( (\alpha, B, \beta) \) is called an immorality if \( B \) contains only one vertex. Frydenberg (1990) notes that \( (\alpha, B, \beta) \) is a minimal complex in \( G \) if and only if \( G_{B \cup \{\alpha, \beta\}} \) looks like the chain graph of Figure 3. For any subset \( A \subseteq V \) such that \( \{\alpha, B, \beta\} \subseteq A \), \( (\alpha, B, \beta) \) is a minimal complex in \( G_A \) if and only if it is a minimal complex in \( G \).

A chain graph \( \tilde{G} \equiv (\tilde{V}, \tilde{E}) \) is larger than the chain graph \( G \equiv (V, E) \), indicated by \( G \subseteq \tilde{G} \), if \( V \subseteq \tilde{V} \) and \( E \subseteq \tilde{E} \). Thus, if \( V = \tilde{V} \) and \( G^u = \tilde{G}^u \), then \( G \subseteq \tilde{G} \) if and only if they differ only in that some directed edges (arrows) in \( G \) may be converted to undirected edges (lines) in \( \tilde{G} \). In this case, each \( \tau \in T(G) \) is contained in a unique \( \tilde{\tau} \equiv \tilde{\psi}(\tau) \in T(\tilde{G}) \), and \( \tilde{\psi} : T(G) \to T(\tilde{G}) \) is an order-preserving mapping, i.e., a poset homeomorphism, under the partial orders induced by \( \leq_G \) and \( \leq_{\tilde{G}} \), respectively. Furthermore, \( \text{bd}_G(\tau) = \text{bd}_{\tilde{G}}(\tilde{\tau}) \) and \( \text{cl}_G(\tau) = \text{cl}_{\tilde{G}}(\tilde{\tau}) \).

The moral graph, \( G^m \), determined by \( G \), is defined to be the undirected graph \( G^m \equiv (V, E^m) \), where \( E^m = E^u \cup_{\tau \in T(G)} E^*(\text{bd}(\tau)) \). That is, \( G^m \) is the skeleton \( G^u \) augmented by all (undirected) edges needed to make \( \text{bd}(\tau) \) complete in \( G \) for every chain component
Figure 3: A simple chain graph. Here \((\alpha, B, \beta)\) is a minimal complex.

\(\tau \in T(G)\). Equivalently, \(G^m\) is obtained from \(G^u\) by adding an undirected edge \(\alpha \leftarrow \beta\) whenever \((\alpha, B, \beta)\) is a minimal complex in \(G\); thus \(G \subseteq G^m\).

For any subset \(A \subseteq V\), \((G_A)^m = (A, (E_A)^m)\), where \((E_A)^m = (E_A)^u \cup \{e_{\tau} \in S(G_A) \mid E^*(bd_A(\tau'))\}\); thus \((G_A)^m \subseteq (G^m)_A\). For every chain component \(\tau \in T(G)\), \((G_{cl(\tau)})^m\) is the skeleton \((G_{cl(\tau)})^u\) augmented by those undirected edges \(\alpha \leftarrow \beta\) such that \((\alpha, B, \beta)\) is a minimal complex for some \(B \subseteq \tau\). Furthermore, \((G_{cl(\tau)})^m = (G^m)_{cl(\tau)}\) for every terminal chain component \(\tau \in T(G)\).

3 The Markov Properties for Chain Graphs

First, we describe informally the Markov properties specified by ADGs and UDGs, then do so formally for chain graphs, which include both ADGs and UDGs as special cases. The Markov property specified by an ADG embodies the hypothesis that for each vertex \(\alpha \in V\), the parents of \(\alpha\) are the only direct influences on \(\alpha\). That is, a probability measure \(P\) on \(\mathcal{X}\) satisfies the (local) Markov property specified by \(G\) if, under \(P\), each \(X_\alpha, \alpha \in V\), is conditionally independent of its nondescendants given its parents, where the *nondescendants* of \(\alpha\) are those \(\beta \in V\) such that there is no directed path from \(\alpha\) to \(\beta\). When \(P\) admits a joint density \(p = p(V)\) with respect to some product measure on \(\mathcal{X}\), this property implies a factorization of \(p\) given by:

\[
(3.1) \quad p(V) = \prod_{\alpha \in V} p(\alpha \mid pa(\alpha)).
\]

The class of probability models which can be defined in this way were introduced by Wermuth and Lauritzen (1983) and Kuiveri et al (1984) and are a subclass of their class of recursive
causal models.

A probability measure $P$ on $\mathcal{X}$ satisfies the (local) Markov property specified by a UDG (see for example Figure 1(b)) if, under $P$, each $X_\alpha, \alpha \in V$ is conditionally independent of all others given its neighbors. If $P$ admits a positive joint density $p$ on $\mathcal{X}$, this (local) Markov property does not imply a simple factorization of $p$, unless the UDG is also decomposable. The simplicity of decomposable models has been exploited in a number of contexts—see for example Lauritzen and Spiegelhalter (1988), Madigan and Mosurski (1990), Dawid and Lauritzen (1993), and Madigan and Raftery (1994). For a more detailed exposition of Markov properties with respect to ADGs and UDGs, we refer the reader to Lauritzen et al. (1990) and Dawid and Lauritzen (1993).

Frydenberg (1990) defined three versions of the Markov property for a general chain graph $G \equiv (V, E)$. First, for three pairwise disjoint subsets $A, B,$ and $C$ of $V$, and $P$ a probability measure on $\mathcal{X}$, we write $A \perp B \mid C[P]$ if $X_A$ and $X_B$ are conditionally independent given $X_C$ under $P$. Trivially, $A \perp B \mid C[P]$ if $A = \emptyset$ or $B = \emptyset$. If $A, B,$ and $C$ are not disjoint, then $A \perp B \mid C[P]$ is defined to mean $[A \setminus (B \cup C)] \perp [B \setminus (A \cup C)] \mid C[P]$. Dawid (1980) showed that the following properties hold for any probability measure $P$:

\begin{itemize}
  \item [(CI1)] $A \perp B \mid C[P]$ implies $B \perp A \mid C[P]$
  \item [(CI2)] $A \perp B \cup C \mid D[P]$ implies $A \perp B \mid D[P]$
  \item [(CI3)] $A \perp B \cup C \mid D[P]$ implies $A \perp B \mid C \cup D[P]$
  \item [(CI4)] $(A \perp B \mid D[P]$ and $A \perp C \mid D \cup B[P])$ implies $A \perp B \cup C \mid D[P]$
\end{itemize}

whenever $A, B, C,$ and $D$ are disjoint subsets of $V$. Note that the converse implication in (CI4) is also valid by (CI2) and (CI3).

**DEFINITION 3.1.** A probability measure $P$ on $\mathcal{X}$ is said to be:

- (P) **Pairwise $G$-Markovian** if $\alpha \perp \beta \mid [V \setminus \phi(\alpha)] \setminus \{\alpha, \beta\}[P]$ whenever $\beta \not\in \phi(\alpha)$ and $\beta$ and $\alpha$ are not adjacent;

- (L) **Local $G$-Markovian** if $\alpha \perp [V \setminus \phi(\alpha)] \setminus \text{cl}(\alpha) \mid \text{bd}(\alpha)[P]$ for all $\alpha$;

- (G) **Global $G$-Markovian** if $A \perp B \mid C[P]$ whenever $C$ separates $A$ and $B$ in $(G_{\text{an}(A \cup B \cup C)})^*.$

Frydenberg (1990, p.334) notes that if $G$ is a UDG, then $P$ is global $G$-Markovian if and only if $A \perp B \mid C[P]$ whenever $C$ separates $A$ and $B$ in $G$. For ADGs, Lauritzen et al. (1990) showed that (G) $\Leftrightarrow$ (L) $\Leftrightarrow$ (P) for any $P$ on $\mathcal{X}$. For general chain graphs, Frydenberg (1990) showed that (G) $\Rightarrow$ (L) $\Rightarrow$ (P) for any $P$ on $\mathcal{X}$, while if $P$ satisfies the following property CI5, then (P), (L), and (G) are equivalent for $P$ (Frydenberg, 1990, Theorem 3.3):
whenever $A, B, C,$ and $D$ are disjoint subsets of $V$. Note that CI5 is satisfied whenever $P$ has a positive joint probability density with respect to some product measure on $\mathcal{X}$. 

**Definition 3.2.** Two chain graphs $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ are called Markov equivalent if, for every product space $\mathcal{X}$ indexed by $V$, the classes of global $G_1$-Markovian and global $G_2$-Markovian probability measures $P$ on $\mathcal{X}$ coincide. In this case we write $G_1 \equiv G_2$.

Theorem 5.6 of Frydenberg (1990) states that two chain graphs are Markov equivalent if and only if they have the same skeleton and the same minimal complexes. However, his treatment of Markov equivalence is entirely restricted to probability measures $P$ satisfying CI5. In the Appendix we remove this restriction and establish the following stronger result, which is of independent interest:

**Theorem 3.1.** Two chain graphs $G_1 \equiv (V, E_1)$ and $G_2 \equiv (V, E_2)$ are Markov equivalent if and only if $G_1$ and $G_2$ have the same skeleton and the same minimal complexes.

**Proof.** See Appendix.

**Remark 3.1.** It follows immediately from Theorem 3.1 that two UDGs are Markov equivalent if and only if they are identical. Thus a UDG is uniquely determined by its associated Markov model. However, the same Markov model may also be specified by other chain graphs, including ADGs (see Corollary 4.5).

**Remark 3.2.** Theorem 3.1 also implies that if a chain graph $G$ is Markov equivalent to some ADG, then $G$ can have no minimal complexes other than immoralities. The converse is false—see Example 4.1 and subsequent discussion.

**Remark 3.3.** If a probability measure has a strictly positive density, $f$, with respect to some product measure $\mathcal{X}$, then $P$ satisfies CI5. That this strict positivity condition on $f$ is not necessary for $P$ to satisfy CI5 is shown by the example of the uniform (with respect to Lebesgue measure on $\mathcal{X} = \mathbb{R}^3$) probability distribution on the skewed cube given by the convex hull of the points $(0,0,0), (1,0,0), (0,1,0), (1,1,0), (0,1,1), (1,1,1), (0,2,1), (1,2,1)$ in $\mathbb{R}^3$. Any random vector in $\mathbb{R}^3$ of the form $(X, g(X), h(X))$, where $g$ and $h$ are strictly increasing functions and $X$ is a non-degenerate random variable, does not satisfy CI5.
4 The Markov Equivalences of a Chain Graph

While our primary result concerns the Markov equivalence of chain graphs and ADGs (Proposition 3.2 below), we also present conditions for Markov equivalence of chain graphs with UDGs (Proposition 4.1) and decomposable UDGs (Proposition 4.3).

PROPOSITION 4.1. Let $G$ be a chain graph. The following are equivalent:

(1) $G$ is Markov equivalent to some (necessarily unique) UDG.

(1') $G$ has no minimal complexes.

(1'') $G \cong G^u$.

PROOF. This follows directly from Theorem 3.1.

EXAMPLE 4.1. Consider the chain graphs $G_1, G_2, G_3,$ and $G_4$ in Figure 4. Since $G_1$ has no minimal complexes, $G_1$ is Markov equivalent to the UDG $G^u \equiv G_2$, by Proposition 4.1. Since $G_3$ has one minimal complex, it is not Markov equivalent to any UDG. Lastly, $G_4$ has one immorality, hence is not Markov equivalent to any UDG.

![Figure 4: Four chain graphs. The first two are Markov equivalent.](image)

In view of Proposition 4.1 and Remark 3.2, one might surmise that a chain graph $G$ is Markov equivalent to some ADG if and only if $G$ has no minimal complexes other than immoralities. Were this so, then every UDG would be Markov equivalent to some ADG, but this is known to be false: only decomposable UDGs are Markov equivalent to some ADG (see Corollary 4.5). Thus, the indecomposable UDG, $G_2$, in Example 4.1 is not Markov equivalent to any ADG; any orientation of the edges of $G_2$ must produce one immorality not
present in $G_2$. Note that $G_2$ has exactly one chain component $\tau$, namely $\tau = G_2$ itself, and this chain component is not decomposable.

Similarly, the chain graph $G_1$ in Example 4.1 is not Markov equivalent to any ADG. (Indeed, $G_1$ is Markov equivalent to $G_2$, and hence to no ADG.) As with $G_2$, $G_1$ has no minimal complexes, but any orientation of its two undirected edges must produce an immorality not present in $G_1$. Note that $G_1$ has exactly two chain components, $\tau_1 \equiv \{\alpha, \beta\}$ and $\tau_2 \equiv \{\gamma, \delta\}$. Looking ahead to Proposition 4.2, we see that the difficulty arises because, for $G = G_1$, the moral graph $(G_{cl(\tau_2)})^m \equiv G_2$ is not decomposable. (Recall that the global Markov property is defined in terms of certain moral graphs.)

Next, the chain graph $G_3$ in Example 4.1 has one minimal complex, which is not an immorality, hence by Remark 3.2, $G_3$ cannot be Markov equivalent to any ADG: either orientation of the single undirected edge in $G_3$ will produce an immorality not present in $G_3$. Here, $G_3$ has three chain components, $\tau_1 = \{\alpha\}$, $\tau_2 = \{\beta\}$, and $\tau_3 = \{\gamma, \delta\}$, and for $G = G_3$, $(G_{cl(\tau_3)})^m \equiv G_2$ is not decomposable.

Finally, the chain graph $G_4$ in Example 4.1 has one minimal complex, an immorality. There are four possible orientations of the two undirected edges of $G_4$ (see Figure 5), of which three produce no additional immoralities, hence produce ADGs to which $G_4$ is Markov equivalent. Here, $G_4$ has two chain components, $\tau_1 = \{\alpha, \beta, \delta\}$ and $\tau_2 = \{\gamma\}$. For $G = G_4$, both $(G_{cl(\tau_1)})^m \equiv G_{\tau_1}$ and $(G_{cl(\tau_2)})^m \equiv G_{\tau_2}$ are decomposable, hence $G_4$ satisfies condition (2') of Proposition 4.2 below, our main result concerning the Markov equivalence of chain graphs and ADGs.

![Figure 5: The four possible orientations of $G_4$. By Theorem 3.1, the first three ADGs are Markov equivalent to $G_4$ while the fourth is not.](image-url)
First we require a lemma:

**Lemma 4.1.** Let $G$ be a chain graph such that $(G_{cl(\tau)})^m$ is decomposable for every chain component $\tau \in \mathcal{T}(G)$. Then $G$ has no minimal complexes other than immoralities.

**Proof.** If $(\alpha, B, \beta)$ is a minimal complex in $G$, then $(\alpha, B, \beta)$ determines a chordless $n$-cycle in $(G_{cl(\tau)})^m$, where $\tau \in \mathcal{T}(G)$ is the unique chain component containing $B$ and where $n = |B| + 2$ (see Figure 3). Thus, if $|B| \geq 2$ then $(G_{cl(\tau)})^m$ is not decomposable.

The converse of Lemma 4.1 is not true: any indecomposable UDG provides a counterexample, for example, $G_2$ in Figure 4. The chain graph $G_1$ also provides a counterexample.

**Proposition 4.2.** Let $G \equiv (V, E)$ be a chain graph. The following are equivalent:

(2) $G$ is Markov equivalent to some (not necessarily unique) ADG, $D$.

(2') $(G_{cl(\tau)})^m$ is decomposable for all $\tau \in \mathcal{T}(G)$.

**Proof.** (2) $\Rightarrow$ (2'): Suppose that $G \sim M D$ but that $(G_{cl(\tau)})^m$ is not decomposable for some $\tau \in \mathcal{T}(G)$. Then $(G_{cl(\tau)})^m$ must contain at least one chordless cycle, $C = \{\alpha_0, \alpha_1, \ldots, \alpha_n \equiv \alpha_0\}$, $n \geq 4$, with $C \subseteq cl(\tau) \equiv bd(\tau) \cup \tau$. First, $C \not\subseteq \tau$; otherwise, any acyclic orientation of the edges of this chordless cycle would create at least one immorality in $D$ that was not present in $G$, hence which was not present in $G$. By Theorem 3.1, this would violate the assumption that $G \not\sim M D$. Thus $|C \cap bd(\tau)| \geq 1$. Also, $|C \cap bd(\tau)| \leq 2$, since bd($\tau$) is complete in $(G_{cl(\tau)})^m$ and $C$ is chordless in $(G_{cl(\tau)})^m$.

If $|C \cap bd(\tau)| = 1$, assume without loss of generality that $C \cap bd(\tau) = \alpha_1$. Because $\alpha_1 \in bd(\tau) = pa(\tau)$ and $\alpha_1$ is adjacent in $G$ to both $\alpha_2 \in \tau$ and to $\alpha_n \in \tau$, necessarily both $\alpha_1 \rightarrow \alpha_2$ and $\alpha_1 \rightarrow \alpha_n$ in $G$ (see Figure 6a). If $|C \cap bd(\tau)| = 2$, assume without loss of generality that $C \cap bd(\tau) = \{\alpha_1, \alpha_2\}$, in which case necessarily both $\alpha_1 \rightarrow \alpha_n$ and $\alpha_2 \rightarrow \alpha_3$ in $G$ (see Figure 6b). Since $C$ is chordless and $n \geq 4$, it follows that any possible orientation of the edges $\alpha_2 \rightarrow \cdot \cdot \cdot \rightarrow \alpha_n$ in the first case, or $\alpha_3 \rightarrow \cdot \cdot \cdot \rightarrow \alpha_n$ in the second case, will create at least one immorality in $D$ that was not present in $G$, again contradicting the assumption that $G \sim M D$.

(2') $\Rightarrow$ (2): Let $G$ be a chain graph satisfying (2'). For each chain component $\tau \in \mathcal{T}(G)$, the edges of the decomposable UDG $(G_{cl(\tau)})^m$ can be oriented by means of the maximum cardinality search (MCS) algorithm to convert $(G_{cl(\tau)})^m$ into a perfect digraph $D(\tau)$, i.e., $D(\tau)$ is an acyclic digraph with no immoralities (see Blair and Peyton (1993, Theorem 2.5)). The MCS algorithm begins by assigning the number 1 to any vertex in $cl(\tau)$, then assigning
the numbers 2, ..., \(|\text{cl}(\tau)|\) consecutively to the remaining vertices in \(\text{cl}(\tau)\), selecting each time the vertex with the most previously numbered neighbors in \((G_{d(\tau)})^m\), breaking ties arbitrarily. Thus we may begin at any vertex in \(\text{bd}(\tau)\), then, since \(\text{bd}(\tau)\) is complete in \((G_{d(\tau)})^m\), number the remaining vertices in \(\text{bd}(\tau)\) before numbering any of the vertices in \(\tau\). With such a numbering of \(\text{cl}(\tau)\), the orientation in \(D^{(\tau)}\) of any edge between a vertex in \(\text{bd}(\tau)\) and a vertex in \(\tau\) agrees with the orientation of that edge in \(G\).

Let \(D\) be the digraph obtained from \(G\) by orienting all undirected edges in \(G\) as follows: for every \(\tau \in T(G)\), orient all edges in \(G_{\tau}\) according to their orientation in \(D^{(\tau)}\). Then \(D\) is acyclic: if \(D\) possesses a directed cycle, \(C\), then \(C\) cannot lie entirely lie within any chain component \(\tau \in T(G)\) (since \(D^{(\tau)}\) is acyclic), hence \(G_{C}\) contains at least one directed edge. But then \(C\) is also a directed cycle in the chain graph \(G\), which is impossible.

Thus, \(D\) is an ADG with the same skeleton as \(G\). To show that \(G \cong D\), by Theorem 3.1 we must show that \(D\) has the same minimal complexes as \(G\). By Lemma 4.1, \(G\) has no minimal complexes other than immoralities. Since any immorality of \(G\) is also an immorality of \(D\), it suffices to show that \(D\) has no immoralities not already in \(G\). Such a new immorality, say \((\alpha, \gamma, \beta)\), must involve at least one undirected edge in \(G\), say \(\alpha \rightarrow \gamma\). (Otherwise, this immorality would also occur in \(G\).) Thus \(\alpha, \gamma \in \tau\) for some \(\tau \in T(G)\), and \(\beta \in \text{cl}(\tau) = \text{bd}(\tau) \cup \tau\). Then \((\alpha, \gamma, \beta)\) is also an immorality in \(D^{(\tau)}\), since \(\alpha\) and \(\beta\) cannot be adjacent in \((G_{d(\tau)})^m\). But \(D^{(\tau)}\) is perfect, so this yields a contradiction, thus completing the proof.

The ADG, \(D\), constructed in the proof of the implication \((2') \Rightarrow (2)\), is obtained from \(G\) by orienting all undirected edges in each induced subgraph \(G_\tau, \tau \in T(G)\), according to
the orientation of these edges in the corresponding ADG, $D^{(r)}$. Thus the orientations of the remaining edges in $D^{(r)}$ are not needed in this construction. The following algorithm makes use of this observation to construct efficiently a Markov equivalent ADG, $D$, from a chain graph $G$ satisfying $(2')$:

The Orientation Algorithm.

For every chain component $\tau \in \mathcal{T}(G)$, perform a modified maximum cardinality search (MCS) on $G_\tau$ as follows:

1. Choose $\alpha_1 = \arg\max_{\alpha \in \tau} \{|pa_G(\alpha)|\}$, breaking ties arbitrarily.

2. For each $i = 2, \ldots, |\tau|$, having chosen $\alpha_1, \ldots, \alpha_{i-1} \in \tau$, now choose:

   $$\alpha_i = \arg\max_{\alpha \in \tau \setminus \{\alpha_1, \ldots, \alpha_{i-1}\}} \{|pa_G(\alpha)| + |\{j \in \{1, \ldots, i-1\} : (\alpha_j, \alpha) \in E\}|\},$$

   breaking ties arbitrarily. This produces an ordering $\{\alpha_1, \ldots, \alpha_{|\tau|}\}$ of all the vertices in $\tau$.

3. Orient the edges of $G_\tau$ in accordance with this ordering, i.e., orient $\alpha_j \rightarrow \alpha_i$ if and only if $(\alpha_j, \alpha_i) \in E$ and $j < i$.

4. All edges in $G$ are now oriented, producing the required ADG, $D$.

In order to verify that a chain graph $G$ does indeed satisfy $(2')$, first apply the Orientation Algorithm to $G$, obtaining for each $\tau \in \mathcal{T}(G)$ an ordering $\alpha_1, \ldots, \alpha_{|\tau|}$ of the vertices in $\tau$. Next, arbitrarily order the vertices in $\text{bd}(\tau)$: $\beta_1, \ldots, \beta_{|\text{bd}(\tau)|}$. Finally apply the “test for zero fill-in” of Tarjan and Yannakakis (1984, p.571) to the reversed ordering $\alpha_{|\tau|}, \ldots, \alpha_1, \beta_{|\text{bd}(\tau)|}, \ldots, \beta_1$ of the vertices of $(G_{d(\tau)})^m$. If this reversed ordering is “zero fill-in” for $(G_{d(\tau)})^m$, then $(G_{d(\tau)})^m$ is chordal, hence decomposable.

Remark 4.1. Once we have constructed one ADG, $D$, that is Markov equivalent to a chain graph $G$ satisfying $(2')$, we can generate all such Markov equivalent ADGs as follows. From $D$, construct the essential graph $D^*$ determined by $D$ by means of the algorithm given by Andersson, Madigan, and Perlman (1995). Here, $D^*$ is the graph obtained from $D$ by converting to undirected edges those directed edges in $D$ that occur with the opposite orientation in at least one ADG that is Markov equivalent to $D$ (hence to $G$). Andersson et al (1995) show that $D^*$ is a chain graph, each of whose chain components is decomposable. Furthermore, they show that all ADGs that are Markov equivalent to $D$ (hence to $G$) can
be obtained by assigning all possible perfect (≡ acyclic and moral) orientations to the edges in these decomposable chain components of \( D^* \).

**Remark 4.2.** An equivalent but more explicit version of condition (2') can be obtained from the following observation. First, for any vertex \( \alpha \in V \) and any subset \( A \subseteq V \), \( \text{ch}_A(\alpha) \equiv \{ \beta \in A \mid (\alpha, \beta) \in E \} \) is the set of all children of \( \alpha \) that occur in \( A \). Then, for any chain component \( \tau \in T(G) \), \((G_{\text{cl}(\tau)})^n\) is decomposable if and only if:

- (i) \( G_\tau \) is decomposable, and
- (ii) For every \( \alpha \in \text{bd}(\tau) \), and every non-adjacent pair \( \gamma, \delta \in \text{ch}_\tau(\alpha), \text{ch}_\tau(\alpha) \setminus \{ \gamma, \delta \} \) separates \( \gamma \) and \( \delta \) in \( G_\tau \) (in particular \( \text{ch}_\tau(\alpha) \setminus \{ \gamma, \delta \} \) must be nonempty), and
- (iii) For every distinct pair \( \alpha, \beta \in \text{bd}(\tau) \), and every pair \( \gamma \in \text{ch}_\tau(\alpha) \setminus \text{ch}_\tau(\beta), \delta \in \text{ch}_\tau(\beta) \setminus \text{ch}_\tau(\alpha), \text{ch}_\tau(\alpha) \cup \text{ch}_\tau(\beta) \setminus \{ \gamma, \delta \} \) separates \( \gamma \) and \( \delta \) in \( G_\tau \) (in particular, \([ \text{ch}_\tau(\alpha) \cup \text{ch}_\tau(\beta) ] \setminus \{ \gamma, \delta \} \) must be nonempty and \( \gamma \) and \( \delta \) must be non-adjacent).

To verify this assertion, first note that the failure of either (i), (ii), or (iii) would imply the existence of a chordless \( n \)-cycle in \((G_{\text{cl}(\tau)})^n\) for some \( n \geq 4 \), hence \((G_{\text{cl}(\tau)})^n\) would be indecomposable. Conversely, if \((G_{\text{cl}(\tau)})^n\) is indecomposable then \((G_{\text{cl}(\tau)})^n\) contains a chordless \( n \)-cycle, \( C, n \geq 4 \). If \( C \subset \tau \), then (i) fails. If \( C \nsubseteq \tau \), proceed as in the proof of the implication \( (2) \Rightarrow (2') \) in Proposition 4.2 to conclude that either (ii) or (iii) must fail (cf., Figure 6).

**Proposition 4.3.** Let \( G \) be a chain graph. The following are equivalent:

1. \( G \) is Markov equivalent to some (necessarily unique) decomposable UDG.
2. \( G \) has no minimal complexes and \( G^n \) is decomposable.
3. \( G \cong G^n \) and \( G^n \) is decomposable.

**Proof.** This follows directly from Proposition 4.1.

By combining Propositions 4.1, 4.2, and 4.3, we can demonstrate in complete generality that the intersection of the classes of graphical Markov models determined by all UDGs and by all ADGs respectively, is the class of graphical Markov models determined by all decomposable UDGs (Corollary 4.5). This will follow directly from Proposition 4.4, whose proof is purely graph-theoretic.

**Proposition 4.4.** Let \( G \) be a chain graph. Then
(1') $G$ has no minimal complexes, and

(2') $(G_{d(\tau)})^m$ is decomposable for all chain components $\tau \in T(G)$, if and only if

(3') $G$ has no minimal complexes and $G^u$ is decomposable.

**Proof.** (3') $\Rightarrow$ (1') and (2'): It suffices to show that $G$ satisfies (2'). Because $G$ has no minimal complexes, $(G_{d(\tau)})^m = (G_{d(\tau)})^u = (G^u)_{d(\tau)}$ for every $\tau \in T(G)$. However, $(G^u)_{d(\tau)}$ is decomposable, since every induced subgraph of a decomposable UDG is decomposable.

(1') and (2') $\Rightarrow$ (3'): It suffices to show that $G^u$ is decomposable. If not, then $G^u$ contains a chordless $n$-cycle, $C = \{\alpha_0, \alpha_1, \ldots, \alpha_n \equiv \alpha_0\}, n \geq 4$. By (2'), $C$ cannot lie entirely within any chain component of $G$, hence $G_C$ must have at least one directed edge. Since $G$ is a chain graph, $G_C$ therefore must have at least two directed edges, say $(\alpha_0, \alpha_1)$ and $(\alpha_k, \alpha_{k-1})$, where $2 \leq k \leq n$, with opposing orientations in $G_C$. Furthermore, we may select these two edges such that either $\alpha_1 = \alpha_{k-1}$ (hence $k = 2$ — see Figure 7(a)) or else $\{\alpha_1, \ldots, \alpha_{k-1}\}$ forms an undirected path in $G_C$ (hence $3 \leq k \leq n$ — see Figure 7(b)). In the first case, $(\alpha_0, \alpha_1, \alpha_2)$ forms an immorality in $G$ (since $C$ is chordless), violating (1'). In the second case, let $\tau \in T(G)$ be the unique chain component of $G$ such that $\{\alpha_1, \ldots, \alpha_{k-1}\} \in \tau$. If $k = n - 1$ or $n$, then $\alpha_0 \in \text{bd}(\tau)$ and $\alpha_k \in \text{bd}(\tau)$, so $C$ is a chordless cycle in $(G^u)_{d(\tau)} = (G_{d(\tau)})^u$, violating (2'). If $3 \leq k \leq n - 2$ (in which case $n \geq 5$), then $\alpha_0 \in \text{bd}(\tau)$ and $\alpha_k \in \text{bd}(\tau)$, but $\alpha_0$ and $\alpha_k$ are not adjacent in $G$. But then $\alpha_0$ and $\alpha_k$ are adjacent in $(G_{d(\tau)})^m$, hence $\{\alpha_0, \alpha_1, \ldots, \alpha_{k+1} \equiv \alpha_0\}$ forms a chordless $(k+1)$-cycle in $(G_{d(\tau)})^m$, again violating (2'). This completes the proof.

**Corollary 4.5.** Let $G$ be a chain graph. Then

1. $G$ is Markov equivalent to some UDG, and
2. $G$ is Markov equivalent to some ADG,

if and only if

3. $G$ is Markov equivalent to some decomposable UDG (namely, $G^u$).

**Proof.** This follows immediately from Propositions 4.1 – 4.4.

It is easily seen from the proof of the implication (1') and (2') $\Rightarrow$ (3') that Proposition 4.4 remains true if (1') is replaced by the weaker condition (1''), as follows:

**Proposition 4.6.** Let $G$ be a chain graph. Then
Figure 7: The induced subgraph $G_C$, where $C$ is a chordless $n$-cycle in $G^u$.

$(1''')$ $G$ has no immoralities, and

$(2')$ $(G_d(T))^m$ is decomposable for all chain components $\tau \in T(G)$,

if and only if

$(3')$ $G$ has no minimal complexes and $G^u$ is decomposable.

Note, however, that Propositions 4.4 and 4.6 are not true if $(3')$ is replaced by the following weaker condition:

$(3''')$ $G$ has no immoralities and $G^u$ is decomposable.

The chain graph $G_3$ in Example 4.1 provides a simple counterexample.

Propositions 4.4 and 4.6 give equivalent graphical characterizations of those chain graphs that are Markov equivalent simultaneously to some UDG and to some ADG. It is natural, then, to seek graphical characterizations of those chain graphs that are not Markov equivalent to any UDG or to any ADG (the shaded region in Figure 2).

It follows immediately from Theorem 3.1 that if $G$ has at least one minimal complex that is not an immorality, then $G$ cannot be Markov equivalent to any UDG or to any ADG: for example, $G_3$ in Figure 4. However, the condition is not necessary: the chain graphs $G_5$ and $G_6$ in Figure 8 have no minimal complexes other than immoralities, yet by Propositions 4.1 and 4.2, are not Markov equivalent to any UDG or ADG. The final result of this section, Proposition 4.7, presents three graphical conditions, $(4') - (4''')$, each of which is both necessary and sufficient for a chain graph not to be Markov equivalent to any UDG or ADG.
Figure 8: Two chain graphs with no minimal complexes other than immoralities, yet which are not Markov equivalent to any UDG or ADG.

**Proposition 4.7.** Let $G$ be a chain graph. The following are equivalent:

(4) $G$ is not Markov equivalent to any UDG or to any ADG.

(4') $G$ has at least one minimal complex, and $(G_{cl(\tau)})^m$ is indecomposable for at least one chain component $\tau \in \mathcal{T}(G)$.

(4'') (a) $G$ has at least one minimal complex that is not an immorality; or

(b) $G$ has at least one immorality, and $(G_{cl(\tau)})^m$ is indecomposable for at least one chain component $\tau \in \mathcal{T}(G)$.

(4'''') (a) $G$ has at least one minimal complex that is not an immorality; or

(b) $G$ has at least one immorality, and for at least one chain component $\tau \in \mathcal{T}(G)$, either

(i) $G_{\tau}$ is indecomposable; or

(ii) There exists an $\alpha \in \text{bd}(\tau)$ and there exists a non-adjacent pair $\gamma, \delta \in \text{ch}_r(\alpha)$, such that there is a path $\pi$ in $\tau$ from $\gamma$ to $\delta$ with $[\pi \setminus \{\gamma, \delta\}] \cap \text{ch}_r(\alpha) = \emptyset$; or

(iii) There exists a distinct pair $\alpha, \beta \in \text{bd}(\tau)$ and there exists a pair $\gamma \in \text{ch}_r(\alpha) \setminus \text{ch}_r(\beta), \delta \in \text{ch}_r(\beta) \setminus \text{ch}_r(\alpha)$, such that there exists a path $\pi$ in $\tau$ from $\gamma$ to $\delta$ with $[\pi \setminus \{\gamma, \delta\}] \cap [\text{ch}_r(\alpha) \cup \text{ch}_r(\beta)] = \emptyset$.

**Proof.** This follows immediately from Propositions 4.1 and 4.2 and Remark 4.2.
Appendix. Proof of Theorem 3.1.

Theorem 3.1 extends the Markov equivalence theorem of Frydenberg (1990, Theorem 5.6) to general probability measures. Frydenberg’s restriction to measures that satisfy CI5 can be removed, provided that Markov equivalence is defined in terms of the *global* Markov property. The outline of our proof will closely follow Frydenberg (1990). First, however, we must extend to general UDGs a result obtained by Dawid and Lauritzen (1993, Proposition B.7) for the special case of decomposable graphs.

**Lemma A.1.** Let \((C, D)\) be a decomposition of an undirected graph \(G = (V, E)\) and let \(P\) be a probability measure on \(\mathcal{X}^\ast = \times_{\alpha \in V} \mathcal{X}_\alpha\). Then \(P\) is global \(G\)-Markovian if and only if

1. \(P_C\) is global \(G_C\)-Markovian,
2. \(P_D\) is global \(G_D\)-Markovian, and
3. \(C \perp D \mid C \cap D[P]\).

**Proof.** The result is easily verified if \(C = V\) or \(D = V\), so we may assume that the decomposition is proper.

First assume that \(P\) is global \(G\)-Markovian. In order to establish (i), it suffices to show that if \(S\) separates \(A\) and \(B\) in \(G_C\), then \(S\) separates \(A\) and \(B\) in \(G\). If the latter fails, there exists a path \(\pi = \{\alpha_0, \ldots, \alpha_n\} \subseteq V \equiv C \cup D\) such that \(\alpha_0 \in A, \alpha_n \in B,\) and \(\pi \cap S = \emptyset\). By assumption, \(\pi \not\subseteq C\), hence \(\alpha_i \in D \setminus C\) for some \(i, 1 \leq i \leq n - 1\) (see Figure 9). Since \(C \cap D\) separates \(C \setminus D\) and \(D \setminus C\), there exist integers \(j\) and \(k\) such that \(0 \leq j \leq i - 1,\) \(i + 1 \leq k \leq n\), and \(\alpha_j, \alpha_k \in C \cap D\). Because \(C \cap D\) is complete, \(\alpha_j\) and \(\alpha_k\) must be adjacent in \(G\), hence \(\pi' \equiv \{\alpha_0, \ldots, \alpha_j, \alpha_k, \ldots, \alpha_n\}\) is a path in \(G_C\) between \(A\) and \(B\) that does not intersect \(S\), contradicting the assumption that \(S\) separates \(A\) and \(B\) in \(G_C\). Thus (i) must hold, and similarly (ii). Lastly, (iii) holds by the definition of a decomposition.

Conversely, suppose that \(P\) satisfies (i), (ii), and (iii). We must show that

\[
(A.1) \quad A \perp B \mid S[P] \text{ whenever } S \text{ separates } A \text{ and } B \text{ in } G.
\]

Partition \(A = A_1 \cup A_2 \cup A_3,\) \(B = B_1 \cup B_2 \cup B_3,\) \(S = S_1 \cup S_2 \cup S_3,\) where \(A_1 = A \cap (C \setminus D),\) \(A_2 = A \cap (C \cap D),\) \(A_3 = A \cap (D \setminus C),\) etc. Note that either \(A_2 = \emptyset\) or \(B_2 = \emptyset\) (or both); otherwise, since \(C \cap D\) is complete, \(A\) and \(B\) could not be separated in \(G\). Without loss of generality, assume that \(A_2 = \emptyset\). Define \(E = (C \cap D) \setminus (B \cup S) \equiv (C \cap D) \setminus (B_2 \cup S_2),\) so \(B_2 \cup S_2 \cup E = C \cap D\) (see Figure 10). By (iii) and CI2,
Figure 9: The path $\pi$ (dotted line) between $A$ and $B$ in $V \equiv C \cup D$.

Figure 10: The partitions of $A, B, \text{ and } S$ when $A_2 = \emptyset$; $E$ (shaded) = $(C \cap D) \setminus (B \cap S)$.
We shall establish (A.1) by considering a series of six cases.

CASE 1: \( A_3 = B_2 = B_3 = \emptyset \), so \( A = A_1 \) and \( B = B_1 \). Here (A.2) takes the form:

\[(A.2) \quad (A_1 \cup B_1 \cup S_1) \nmid (B_3 \cup S_3) \mid B_2 \cup S_2 \cup E[P].\]

By assumption, \( S \equiv S_1 \cup S_2 \cup S_3 \) separates \( A_1 \) and \( B_1 \) in \( G \), hence \( S_1 \cup S_2 \) separates \( A_1 \) and \( B_1 \) in \( G_C \). In addition, either (a) \( S_1 \cup S_2 \) separates \( A_1 \) and \( E \) in \( G_C \), or (b) \( S_1 \cup S_2 \) separates \( B_1 \) and \( E \) in \( G_C \) (or both). If (a) holds, it follows from (i) that

\[(A.4) \quad A_1 \nmid (B_1 \cup E) \mid S_1 \cup S_2[P].\]

Also, by (A.3) and CI3,

\[(A.5) \quad A_1 \nmid S_3 \mid B_1 \cup S_1 \cup S_2 \cup E[P].\]

By CI4, (A.4) and (A.5) combine to yield

\[(A.6) \quad A_1 \nmid (B_1 \cup S_3 \cup E) \mid S_1 \cup S_2[P],\]

which, by CI2 and CI3, implies that

\[(A.7) \quad A_1 \nmid B_1 \mid S_1 \cup S_2 \cup S_3[P];\]

this is (A.1) in Case 1. If (b) holds, we again obtain (A.7) by reversing the roles of \( A_1 \) and \( B_1 \).

CASE 2: \( A_3 = B_1 = B_3 = \emptyset \), so \( A = A_1 \) and \( B = B_2 \). Here (A.2) takes the form:

\[(A.8) \quad (A_1 \cup S_1) \nmid S_3 \mid B_2 \cup S_2 \cup E[P].\]

By assumption, \( S \equiv S_1 \cup S_2 \cup S_3 \) separates \( A_1 \) and \( B_2 \) in \( G \), hence \( S_1 \cup S_2 \) separates \( A_1 \) and \( B_2 \) in \( G_C \). Furthermore, since \( C \cap D \) is complete, \( S_1 \cup S_2 \) separates \( A_1 \) and \( E \) in \( G_C \). It follows from (i) that

\[(A.9) \quad A_1 \nmid (B_2 \cup E) \mid S_1 \cup S_2[P].\]

Also, by (A.8) and CI3,

\[(A.10) \quad A_1 \nmid S_3 \mid B_2 \cup S_1 \cup S_2 \cup E[P],\]

which combines with (A.9) to yield

\[(A.11) \quad A_1 \nmid (B_2 \cup S_3 \cup E) \mid S_1 \cup S_2[P].\]

Therefore, by CI2 and CI3, we obtain

\[(A.12) \quad A_1 \nmid B_2 \mid S_1 \cup S_2 \cup S_3[P],\]

\[22\]
which is (A.1) in Case 2.

**Case 3:** \( A_3 = B_1 = B_2 = \emptyset \), so \( A = A_1 \) and \( B = B_3 \). Now (A.2) becomes

\[
(A.13) \quad (A_1 \cup S_1) \perp (B_3 \cup S_3) \mid S_2 \cup E[P].
\]

By assumption, \( S \equiv S_1 \cup S_2 \cup S_3 \) separates \( A_1 \) and \( B_3 \) in \( G \). Also, either (a) \( S_1 \cup S_2 \) separates \( A_1 \) and \( E \) in \( G_C \), or (b) \( S_2 \cup S_3 \) separates \( B_3 \) and \( E \) in \( G_D \) (or both). If (a) holds, it follows from (i) that

\[
(A.14) \quad A_1 \perp E \mid S_1 \cup S_2[P].
\]

By (A.13) and CI3,

\[
(A.15) \quad A_1 \perp (B_3 \cup S_3) \mid S_1 \cup S_2 \cup E[P].
\]

But (A.14) and (A.15) combine by CI4 to yield

\[
(A.16) \quad A_1 \perp (B_3 \cup S_3 \cup E) \mid S_1 \cup S_2[P].
\]

Thus, by CI2 and CI3, we obtain

\[
(A.17) \quad A_1 \perp B_3 \mid S_1 \cup S_2 \cup S_3[P],
\]

which is (A.1) in Case 3. If (b) holds, (A.17) is obtained by reversing the roles of \( (A_1, C) \) and \( (B_3, D) \) and replacing (i) by (ii).

**Case 4:** \( A_3 = \emptyset \) (so \( A = A_1 \)) and \( B \) arbitrary (so \( B = B_1 \cup B_2 \cup B_3 \)). We will establish (A.1) by combining Cases 1, 2, and 3. By assumption, \( S \) separates \( A_1 \) and \( B_1 \cup B_2 \cup B_3 \) in \( G \); hence \( S' \equiv B_2 \cup S \) separates \( A_1 \) and \( B_1 \) in \( G \). Thus, by Case 1,

\[
(A.18) \quad A_1 \perp B_1 \mid B_2 \cup S[P].
\]

Also, \( S \) separates \( A_1 \) and \( B_2 \) in \( G \), so Case 2 applies to yield

\[
(A.19) \quad A_1 \perp B_2 \mid S[P].
\]

Then (A.18) and (A.19) combine by CI4 to give

\[
(A.20) \quad A_1 \perp (B_1 \cup B_2) \mid S[P].
\]

Next, since \( B_3 \cup S \) also separates \( A_1 \) and \( B_1 \cup B_2 \) in \( G \), (A.20) yields

\[
(A.21) \quad A_1 \perp (B_1 \cup B_2) \mid B_3 \cup S[P].
\]

However, by Case 3,

\[
(A.22) \quad A_1 \perp B_3 \mid S[P],
\]

which combines with (A.21) via CI4 to yield

23
(A.24) \[ A_1 \ni B \mid A_3 \cup S[P]. \]

Next, by Case 5,

(A.25) \[ A_3 \ni B \mid S[P]. \]

Then, (A.24) and (A.25) combine via Cl4 to yield

(A.26) \[(A_1 \cup A_3) \ni B \mid S[P],\]

which is (A.1) in Case 6. This completes the proof of Lemma A.1.

Lemma A.2 (cf. Frydenberg (1990), Proposition 2.2). Let \( G \equiv (V, E) \) be an undirected graph, \( P \) a probability measure on \( \mathcal{A} \), and \( C \) a proper subset of \( V \). If \( C \) is simplicial in \( G \) then the following are equivalent:

(i) \( P \) is global \( G \)-Markovian;

(ii) \( P_{V \setminus C} \) is global \( G_{V \setminus C} \)-Markovian, \( P_{\text{d}(C)} \) is global \( G_{\text{d}(C)} \)-Markovian, and \((V \setminus \text{cl}(C)) \ni \text{bd}(C)[P] .\)

Proof. This follows from Lemma A.1, since \((V \setminus C, \text{cl}(C))\) forms a decomposition of \( G \) with \((V \setminus C) \cap \text{cl}(C) = \text{bd}(C) .\)

Lemma A.3 (cf. Frydenberg (1990), Lemma 3.1). If \( \tau \) is a terminal chain component in a chain graph \( G \equiv (V, E) \) and \( P \) is global \( G \)-Markovian, then \( P_{V \setminus \tau} \) is global \( G_{V \setminus \tau} \)-Markovian.

Proof. We must show that \( A \ni B \mid S[P] \), if \( A, B, \) and \( S \) are subsets of \( V \setminus \tau \) such that \( S \) separates \( A \) from \( B \) in \((G_{V \setminus \tau})_{\text{an}(A \cup B \cup S)}^m \equiv (G_{\text{an}(A \cup B \cup S)})^m \). Because \( V \setminus \tau \) is a \( G \)-anterior set, however, \( \text{an}(A \cup B \cup S) = \text{an}(A \cup B \cup S) \), so \( S \) separates \( A \) from \( B \) in \((G_{\text{an}(A \cup B \cup S)})^m \). Since \( P \) is global \( G \)-Markovian, \( A \ni B \mid S \).

Lemma A.4 (cf. Frydenberg (1990), Lemma 3.2). If \( G \equiv (V, E) \) is a chain graph and \( P \) is global \( G \)-Markovian, then \( P \) is global \( G^m \)-Markovian.
PROOF. Since $G^m$ is an undirected graph, it suffices to show that $A \parallel B \mid S[P]$ whenever $S$ separates $A$ from $B$ in $G^m$. However, $(G_{an(A \cup B \cup S)})^m \subseteq G^m$, so $S$ also separates $A$ from $B$ in $(G_{an(A \cup B \cup S)})^m$ and the result follows.

LEMMA A.5 (cf. Frydenberg (1990), Theorem 3.3). Let $G \equiv (V, E)$ be a chain graph and $P$ a probability measure on $X$. Then the following are equivalent:

(i) $P$ is global $G$-Markovian;

(ii) $P_A$ is global $(G_A)^m$-Markovian for all $G$-anterior subsets $A \subseteq V$.

PROOF. To prove that (i) $\Rightarrow$ (ii), suppose that $P$ is global $G$-Markovian and $A$ is a $G$-anterior set. Because $A$ is obtained from $V$ by stepwise removal of terminal chain components, it follows from Lemma A.3 that $P_A$ is global $G_A$-Markovian. By Lemma A.4, $P_A$ is global $(G_A)^m$-Markovian. The implication (ii) $\Rightarrow$ (i) is immediate from the definition of the global Markov property and its restatement for undirected graphs immediately following Definition 3.1.

PROPOSITION A.1 (cf. Frydenberg (1990), Corollary 3.4). Let $G \equiv (V, E)$ be a chain graph, $\tau$ a terminal chain component, and $P$ a probability measure on $X$. The following are equivalent:

(i) $P$ is global $G$-Markovian;

(ii) $P_{V \setminus \tau}$ is global $G_{V \setminus \tau}$-Markovian and $P$ is $G^m$-Markovian;

(iii) $P_{V \setminus \tau}$ is global $G_{V \setminus \tau}$-Markovian, $P_{cl(\tau)}$ is global $(G_{cl(\tau)})^m$-Markovian, and $\tau \parallel (V \setminus cl(\tau)) \mid bd(\tau)[P]$.

PROOF. That (i) implies (ii) follows from Lemmas A.3 and A.4. That (ii) implies (iii) follows from Lemma A.2 (applied to $G^m$) and the facts that $\tau$ is simplicial in $G^m$ and that $(G^m)_{cl(\tau)} = (G_{cl(\tau)})^m$.

Assume now that (iii) holds. To verify (i), by Lemma A.5 it suffices to show that $P_A$ is global $(G_A)^m$-Markovian for all $G$-anterior sets $A \subseteq V$. If $A \subseteq V \setminus \tau$, then $A$ is $G_{V \setminus \tau}$-anterior, so by Lemma A.5 applied to $G_{V \setminus \tau}$, $P_A$ is global $(G_A)^m$-Markovian. If $A \not\subseteq V \setminus \tau$, then $\tau \not\subseteq A$. We shall apply Lemma A.2 with $(G, C, P)$ replaced by $((G_A)^m, \tau, P_A)$ to conclude that $P_A$ is global $(G_A)^m$-Markovian. In order that Lemma A.2 be applicable, we must verify the following four facts:
(a) \( \tau \) is simplicial in \((G_A)^m\): Since \( \tau \) is terminal in \( G \), it is terminal in \( G_A \). Therefore the boundary of \( \tau \) with respect to \((G_A)^m\) is \( bd_A(\tau) \), which is complete in \((G_A)^m\).

(b) \( P_{A\setminus \tau} \) is global \(((G_A)^m)_{A\setminus \tau}\)-Markovian: It is straightforward to verify that \( A \setminus \tau \) is a \( G_{V \setminus \tau} \)-anterior subset of \( V \setminus \tau \), so by Lemma A.5 applied to \( G_{V \setminus \tau} \) and \( P_{V \setminus \tau} \), \( P_{A\setminus \tau} \) is global \(((G_A)^m)_{A\setminus \tau}\)-Markovian. Since \( ((G_A)^m)_{A\setminus \tau} \supseteq ((G_A)_{A\setminus \tau})^m = (G_{A\setminus \tau})^m \), (b) holds.

(c) \( P_{cl_A(\tau)} \) is global \(((G_A)^m)_{cl_A(\tau)}\)-Markovian: Since \( A \) is \( G \)-anterior, \( cl_A(\tau) = cl(\tau) \). Also, \(((G_A)^m)_{cl(\tau)} = ((G_A)_{cl(\tau)})^m = (G_{cl(\tau)})^m \), so (c) follows from the second relation in (iii).

(d) \( \tau \parallel (A \setminus cl_A(\tau)) \mid bd_A(\tau)[P] \): This follows from the third relation in (iii) and CI2, since \( bd_A(\tau) = bd(\tau) \) and \( cl_A(\tau) = cl(\tau) \).

**Lemma A.6** (cf. Frydenberg (1990), Lemma 3.1). Let \( \tilde{\tau} \) be a terminal chain component of a chain graph \( \tilde{G} \equiv (V, \tilde{E}) \) and let \( \tau \) be a connected subset of \( \tilde{\tau} \). Let \( G \equiv (V, E) \) be the chain graph which differs from \( \tilde{G} \) only in that all edges in \( \tilde{G} \) between \( \tilde{\tau} \setminus \tau \) and \( \tau \) are changed into arrows towards \( \tau \) in \( G \). If \( \tau \) is simplicial in \( \tilde{G}^m \), or equivalently, if \( (\tilde{G}_{cl(\tau)})^m = (G_{cl(\tau)})^m \), then \( G \) and \( \tilde{G} \) are Markov equivalent.

**Proof.** First recall that \( bd(\tau) \) and \( cl(\tau) \) are the same relative to \( G \) and to \( \tilde{G} \). Then the proof of the result is obtained from that of Lemma 5.1 of Frydenberg (1990) with the following minor changes: replace “Markovian” by “global Markovian”, “Corollary 3.4” by “Proposition A.1 (the equivalence of (i) and (ii))”, and “Proposition 2.2” by “Lemma A.2”.

[In the fourth line of Frydenberg’s proof, “\( G \)” should be “\( \tilde{G}^m \)”.

**Proposition A.2** (cf. Frydenberg (1990), Proposition 5.2). Let \( G \equiv (V, E) \) and \( \tilde{G} \equiv (V, \tilde{E}) \) be two chain graphs such that \( G^m = \tilde{G}^m \) and \( G \subseteq \tilde{G} \), i.e., \( \tilde{G} \) might have lines where \( G \) has arrows. If each chain component \( \tau \in T(G) \) is simplicial in \((\tilde{G}_{cl(\tau)})^m\), or equivalently, if \( (\tilde{G}_{cl(\tau)})^m = (G_{cl(\tau)})^m \), then \( G \) and \( \tilde{G} \) are Markov equivalent.

**Proof.** The proof of this result follows that of Proposition 5.2 of Frydenberg (1990) if “Markovian” is replaced by “global Markovian”, “Lemma 5.1” by “Lemma A.6”, and “Corollary 3.4” by “Proposition A.1 (the equivalence of (i) and (iii))”. [In the eighth line of Frydenberg’s proof, “\( P \)” should be “\( P_{W(\tau)} \)” in the ninth line, “\( P_{cl(\tau)} \)” should be “\( P_{cl(\tau)} \)”.

**Lemma A.7** (cf. Frydenberg (1990), p.346). Let \( G \equiv (V, E) \) and \( \tilde{G} \equiv (V, \tilde{E}) \) be two chain graphs such that \( G^m = \tilde{G}^m \) and \( G \subseteq \tilde{G} \). Then \( G \) and \( \tilde{G} \) have the same minimal complexes if and only if any minimal complex in \( G \) is a minimal complex in \( \tilde{G} \).

**Proof.** “only if” is trivial. “if”: Let \( (\alpha, B, \beta) \) be a minimal complex in \( \tilde{G} \). Note that \( G_{(\alpha, \beta) \cup B} \) and \( \tilde{G}_{(\alpha, \beta) \cup B} \) may differ only in that some lines in \( \tilde{G}_B \) might occur as arrows in \( G_B \).

26
But any occurrence of arrows in $G_B$ would create at least one minimal complex in $G_{(\alpha,\beta) \cup B}$ which, by hypothesis, must also exist in $\tilde{G}_{(\alpha,\beta) \cup B}$, violating the fact that $\tilde{G}_B$ has only lines, not arrows. Thus, $G_{(\alpha,\beta) \cup B}$ and $\tilde{G}_{(\alpha,\beta) \cup B}$ are identical, hence $\{\alpha, B, \beta\}$ is also a minimal complex in $G_{(\alpha,\beta) \cup B}$.

**Lemma A.8** (≡ Proposition 5.3 of Frydenberg (1990)). Let $G \equiv (V, E)$ and $\tilde{G} \equiv (V, \tilde{E})$ be two chain graphs such that $G^u = \tilde{G}^u$ and $G \subseteq \tilde{G}$. The following are equivalent:

(i) Each chain component $\tau \in T(G)$ is simplicial in $(\tilde{G}_{cl(\tau)})^m$;

(ii) $G$ and $\tilde{G}$ have the same minimal complexes.

**Proof.** Apply Lemma A.7 – see Frydenberg (1990), pp. 346–347.

**Proposition A.3** (cf. Frydenberg (1990), p.347). Let $G \equiv (V, E)$ and $\tilde{G} \equiv (V, \tilde{E})$ be two chain graphs such that $G^u = \tilde{G}^u$ and $G \subseteq \tilde{G}$. If $G$ and $\tilde{G}$ have the same minimal complexes, then $G$ and $\tilde{G}$ are Markov equivalent.

**Proof.** This follows immediately from Proposition A.2 and Lemma A.8.

If $G \equiv (V, E)$ and $\tilde{G} \equiv (V, \tilde{E})$ are two chain graphs such that $G$ and $\tilde{G}$ have the same vertex set, define $G \cup \tilde{G} = (V, E \cup \tilde{E})$. Let $G \vee \tilde{G}$ be the graph obtained from $G \cup \tilde{G}$ by changing into a line any arrow that is part of a directed cycle in $G \cup \tilde{G}$. Frydenberg (1990, p.347) notes that $G \vee \tilde{G}$ is a chain graph; clearly $G \vee \tilde{G}$ is the (unique) smallest chain graph larger than both $G$ and $\tilde{G}$.

**Proposition A.4** (≡ Proposition 5.4 of Frydenberg (1990)). Let $G \equiv (V, E)$ and $\tilde{G} \equiv (V, \tilde{E})$ be two chain graphs such that $G^u = \tilde{G}^u$. Then $G$ and $\tilde{G}$ have the same minimal complexes if and only if they both have the same minimal complexes as $G \vee \tilde{G}$.

Our main result is now at hand:

**Theorem 3.1** (cf. Frydenberg (1990), Theorem 5.6). Two chain graphs $G \equiv (V, E)$ and $\tilde{G} \equiv (V, \tilde{E})$ are Markov equivalent if and only if $G^u = \tilde{G}^u$ and $G$ and $\tilde{G}$ have the same minimal complexes.

**Proof.** Since $G \vee \tilde{G}$ is larger than both $G$ and $\tilde{G}$, the "if" assertion is a direct consequence of Propositions A.3 and A.4. The "only if" assertion is proved by Frydenberg (1990, pp.347–348).

**Corollary A.1** (cf. Frydenberg (1990), Proposition 5.7). For any chain graph $G$, there exists a unique largest chain graph that is Markov equivalent to $G$.

**Proof.** This follows from Proposition A.4 and Theorem 3.1 – see Frydenberg.
Remark A.1. Because the global, local, and pairwise Markov properties coincide for probability measures that satisfy CI5, Frydenberg's Markov equivalence theorem (1990, Theorem 5.6) is valid regardless of whether Markov equivalence is defined in terms of the global, local, or pairwise property. This implies that the "only if" assertion in our Theorem 3.1 also remains valid if we define Markov equivalence in terms of the local or pairwise, rather than global, Markov property. However, the "if" assertion in Theorem 3.1 does not remain valid in this case. Consider, for example, the pair of graphs $G_7$ and $G_8$ in Figure 11. Because $G_8$ is an ADG, its global, local, and pairwise Markov properties coincide; they are given by the single condition 12 $\not\perp\!\!\!\perp$ 3 4 $\mid$ 5. For the UDG $G_7$, the global Markov property is also given by 12 $\not\perp\!\!\!\perp$ 3 4 $\mid$ 5, whereas its local and pairwise Markov properties are the following:

- **local:** $1 \not\perp\!\!\!\perp 34 \mid 25, \ 2 \not\perp\!\!\!\perp 34 \mid 15, \ 3 \not\perp\!\!\!\perp 12 \mid 45, \ 4 \not\perp\!\!\!\perp 12 \mid 35;$
- **pairwise:** $1 \not\perp\!\!\!\perp 245, \ 1 \not\perp\!\!\!\perp 34 \mid 235, \ 2 \not\perp\!\!\!\perp 3 \mid 145, \ 2 \not\perp\!\!\!\perp 4 \mid 135.$

If $X$ and $Y$ are independent and non-degenerate random variables, then $(X_1, X_2, X_3, X_4, X_5) \equiv (X, X, X, X, Y)$ satisfies the local and pairwise Markov properties for $G_7$ but not for $G_8$. However, $G_7^u = G_8^u$ and $G_7$ and $G_8$ have the same (\equiv no) minimal complexes. [The construction of such examples is facilitated by Propositions 1 and 2 of Matúš (1992).]

Figure 11: Two chain graphs with identical global Markov properties but non-identical local and pairwise Markov properties.

References


